



ARTICLE

Electrophysiological signatures of reward learning in the rodent touchscreen-based Probabilistic Reward Task

Ann M. Iturra-Mena^{1,2,3}, Brian D. Kangas¹, Oanh T. Luc¹, David Potter¹ and Diego A. Pizzagalli¹✉

© The Author(s), under exclusive licence to American College of Neuropsychopharmacology 2023

Blunted reward learning and reward-related activation within the corticostriatal-midbrain circuitry have been implicated in the pathophysiology of anhedonia and depression. Unfortunately, the search for more efficacious interventions for anhedonic behaviors has been hampered by the use of vastly different preclinical and clinical assays. In a first step in addressing this gap, in the current study, we used event-related potentials and spectral analyses in conjunction with a touchscreen version of the rodent Probabilistic Reward Task (PRT) to identify the electrophysiological signatures of reward learning in rats. We trained 11 rats (5 females and 6 males) on the rodent touchscreen-based PRT and subsequently implanted them with deep electrodes in the anterior cingulate cortex (ACC) and nucleus accumbens (NAc) for local field potentials recordings during the PRT. Behaviorally, the expected responsivity-to-reward profile was observed. At the electrophysiological level, we identified a negative amplitude deflection 250–500 ms after feedback in the ACC and NAc electrodes, as well as power increase in feedback-locked delta (1–5 Hz) and alpha/beta (9–17 Hz) bands in both electrodes for rewarded trials. Using a reverse-translational approach, we identified electrophysiological signatures of reward learning in rats similar to those described in humans. These findings and approaches might provide a useful translational platform to efficiently evaluate novel therapeutics targeting anhedonia.

Neuropsychopharmacology (2023) 48:700–709; <https://doi.org/10.1038/s41386-023-01532-4>

INTRODUCTION

Reward-related processes have long been considered in the study of psychiatric disorders. In particular, impairments in reward learning (i.e., the ability to modulate behavior as a function of reinforcement history) have been associated with major depressive disorder (MDD), and have been found to predict poor disease outcome and treatment response [1–3]. Accordingly, reward learning is a critical domain for the study of mood disorders and target for translational work.

From a behavioral perspective, reward learning can be objectively evaluated using the Probabilistic Reward Task (PRT) [4], which is a recommended assay to probe the Positive Valence Systems in the latest revision of the RDoC matrix [5]. Rooted in signal-detection theory [6], the PRT exposes subjects during discrimination learning to an asymmetric probabilistic reinforcement schedule which, in healthy subjects, reliably evokes a response bias toward the more frequently rewarded stimulus (the so-called “rich stimulus”). Relative to healthy controls, individuals with MDD—and specifically those reporting high levels of anhedonia (loss of pleasure to previously rewarding events)—are characterized by a blunted response bias toward the more frequently rewarded stimulus, thus evincing decreased responsiveness to reward [3, 7, 8]. Critically, and highlighting high translational relevance, similar effects have been repeatedly observed using back-translated versions of the PRT designed for laboratory animals (e.g., rats,

non-human primates) while investigating anhedonic phenotypes [9–13].

Regarding the neural basis of reward learning, studies have emphasized the role of the corticostriatal-midbrain dopaminergic circuitry [2, 14]. In particular, nonhuman primate studies have highlighted dopaminergic “teaching” signals in the form of “reward prediction errors” (which refer to the differences between predicted and received rewards), with positive and negative prediction errors eliciting increased and decreased, respectively, phasic neuronal activity of midbrain dopaminergic neurons that support adaptive learning [15–17].

These neural mechanisms have been indirectly assessed in humans using electrophysiological indexes such as the feedback-related negativity (FRN) [18], which more recently has been conceptualized as reward positivity (RewP) [19–21], because the apparent “negativity” actually reflects a reward-related positivity that is reduced in response to non-rewards [20] (thus, thereafter, we will use the term “FRN/RewP”). This ERP component is thought to be evoked by performance feedback and is usually measured ~200–300 ms post-feedback at fronto-central electrodes. The FRN/RewP is assumed to reflect transmission of dopaminergic signals from striatal regions to the dorsal anterior cingulate cortex (dACC) [22, 23]. Directly relevant here, a prior human PRT study found that pharmacologically-induced reduction in reward learning was associated with a larger (i.e., more negative FRN)

¹Harvard Medical School, McLean Hospital, Belmont, MA 02478, USA. ²Department of Child and Adolescent Psychiatry, Columbia University, New York, NY 10032, USA. ³Data Science Institute, Columbia University, New York, NY 10027, USA. ✉email: dap@mclean.harvard.edu

Received: 2 October 2022 Revised: 18 December 2022 Accepted: 31 December 2022

Published online: 16 January 2023

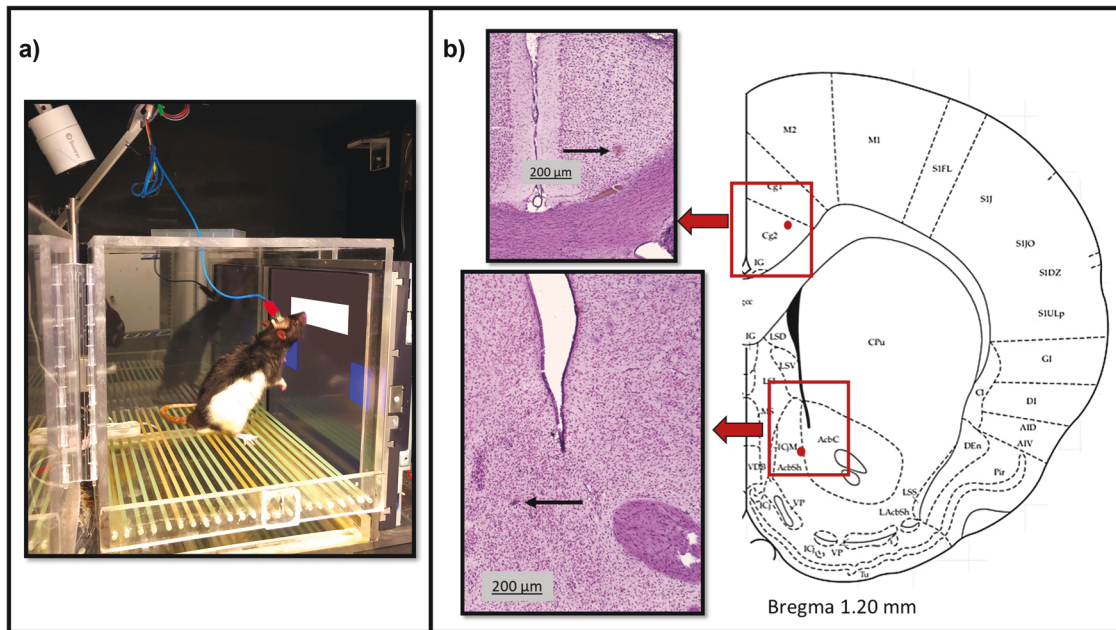


Fig. 1 **Recording setup and electrode placement.** **a** Photograph of the behavioral/recording setup (the rat in the picture is only an example as we used a different breed for these experiments). **b** Representative histology image showing the electrode placement in ACC (upper panel) and NAc (lower panel).

and/or reduced RewP in response to the more frequently rewarded (rich) stimulus [24]. Similarly, among healthy controls, those with the largest reward learning in the PRT were characterized by significantly smaller FRN (i.e., more positive RewP) in response to rich reward feedback relative to participants learning less [23]. Finally, in the latter study, a positive correlation between FRN in response to rich reward feedback and reward learning emerged: thus, a smaller (less negative) FRN correlated with larger reward learning. In the current study, correlational analyses were run between FRN/RewP and reward learning to evaluate whether similar brain-behavior relationships emerged in rats.

In addition to the FRN/RewP, other EEG parameters, including task-induced spectral (oscillatory) changes have been related to different aspects of reward processing [20, 21]. In particular, delta activity has been associated with positive prediction errors (with more delta power when better than expected outcomes are delivered) [25]. Although less studied, feedback-related beta oscillations have also been found to be sensitive to the receipt of unexpected gains and they have been hypothesized to relate to dopaminergic activation after a salient event [26, 27]. Conceptually, in the PRT, the largest positive reward prediction errors is elicited when the least frequently rewarded stimulus (the so-called “lean stimulus”) is rewarded; accordingly, correlational analyses evaluated whether delta and beta oscillations in response to rewarded lean stimuli were particularly associated with higher reward learning.

The overarching aim of the current study was to identify electrophysiological signatures of reward learning using both ERP and EEG spectral analyses while rats performed a touchscreen-based version of the PRT that is functionally identical to the human version [9]. Owing to human PRT findings linking response bias toward the more frequently rewarded stimulus to reward-related activation in the nucleus accumbens (NAc) and dACC [23, 24], electrophysiological signals in the homologous regions of the rat were probed. We hypothesized that the current touchscreen-based rodent version of the PRT would elicit (1) a robust response bias, (2) a FRN/RewP-like waveform, (3) and feedback-locked spectral changes in oscillatory activity in delta and beta frequency bands in

response to reward, similarly to what has been observed in humans.

MATERIALS AND METHODS

Subjects

Eleven Sprague-Dawley rats (6 males weighing 300–350 g and 5 females weighing 250–300 g) were purchased from Charles River Laboratories (Wilmington, MA) and maintained on a 12 h light/dark cycle (lights on 7:00 a.m. to 7:00 p.m.). Rats were initially housed in groups of three and then singly housed following electrode implantation surgery. Access to water was unrestricted and access to food was restricted to daily post-session portions of ~10 to 15 g of rodent chow per subject. All procedures were approved by the McLean Hospital Institutional Animal Care and Use Committee and consistent with the 2010 National Institutes of Health Guide for the Care and Use of Laboratory Animals.

The sample size was calculated using the software G*Power version 3.1.9.7 (Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany) for within factors repeated measures ANOVA and assuming a level of significance of 0.05, an effect size of 0.95 and power of 0.80.

Apparatus

The rat touch-sensitive experimental chamber (Fig. 1a) used in these experiments have been previously described [27]. A chamber made of plexiglas (25 × 30 × 35 cm) was situated in a sound- and light-attenuating enclosure (40 × 60 × 45 cm), with a 17" touch-sensitive screen (1739L, ELO TouchSystems, Menlo Park, CA) comprising the inside right-hand wall of the chamber. Outside of the enclosure, an infusion pump (PHM-100-5, Med Associates, St. Albans, VT) was used to deliver sweetened condensed milk solution (Sysco Corporation, Houston, TX) into the reservoir (diameter: 3 cm) of a custom-designed aluminum receptacle (4 × 5 × 1 cm) mounted 2 cm above the floor and centered on the left-hand inside wall. Above the touchscreen, a speaker bar (NQ576AT, Hewlett-Packard, Palo Alto, CA) was mounted to emit audible feedback. All experimental events and data collection were programmed in E-Prime Professional 2.0 (Psychology Software Tools, Inc., Sharpsburg, PA).

Experimental design and statistical analysis

Probabilistic Reward Task (PRT). Empirical validation and task optimization of the touchscreen-based rat PRT has been previously described elsewhere [9].

Line length discrimination training: Trials began with presentation of a white line on a black background, with its lower edge presented 3 cm above 5 × 5 cm left and right blue response boxes. The length of the line was either 600 × 120 px (31.5 × 6.5 cm: long line) or 200 × 60 px (10.5 × 3.25 cm: short line). Long and short line-length trial types varied in a quasi-random manner across 100-trial sessions such that there were exactly 50 trials of each type, but a given trial type would not be presented more than 5 times in a row. Subjects were differentially reinforced to respond to the left or right response box depending on the length of the white line (e.g., long line: respond left, short line: respond right, or vice versa). Long and short response box designation was counterbalanced across subjects. During line length discrimination training, each correct response was reinforced with 0.1 ml of 30% sweetened condensed milk, paired with an 880-ms yellow screen flash and a 440 Hz tone, and followed by a 5-s blackout period, whereas each incorrect response immediately resulted in a 10-s blackout period. A correction procedure was implemented during initial discrimination training in which each incorrect trial was repeated until a correct response was made [28] and was discontinued after fewer than ten repeats of each trial type occurred in two consecutive sessions. Discrimination training sessions continued without correction until accuracies for both line length trial types were ≥80% correct for two consecutive sessions, concordant with the performance criteria in previous human PRT studies [4, 8, 29]. After this criterion was met, PRT testing commenced.

PRT testing: Subjects were exposed to a 300-trial test session using 3:1 probabilistic reinforcement contingencies, such that a correct response to one of the line lengths (long or short) was reinforced 60% of the time (rich stimulus), whereas a correct response to the other line length was reinforced 20% of the time (lean stimulus). Incorrect responses were never reinforced. The line length associated with the rich and lean contingency was determined for each subject during their final two-line length discrimination training sessions by examining their accuracies and designating the line length with a higher mean accuracy as the stimulus to be rewarded on the lean schedule. This approach was expressly designed to examine response bias generated by responsivity to asymmetrical probabilistic contingencies, rather than the amplification of a preexisting inherent bias that is a function of uncontrolled variables.

PRT data analysis: The implementation of probabilistic contingencies yields two primary dependent measures: response bias and discriminability, which can be quantified using equations derived from signal-detection theory [6, 30, 31] by examining the number of correct and incorrect responses for Rich and Lean trial types. Response Bias is calculated using the following log *b* equation:

$$\log b = 0.5 \times \log \left(\frac{(\text{Rich}_{\text{Correct}} + 0.5) \times (\text{Lean}_{\text{Incorrect}} + 0.5)}{(\text{Rich}_{\text{Incorrect}} + 0.5) \times (\text{Lean}_{\text{Correct}} + 0.5)} \right)$$

High bias values are produced by high numbers of correct responses for rich trials and incorrect responses for lean trials. Discriminability is calculated using the following log *d* equation:

$$\log d = 0.5 \times \log \left(\frac{(\text{Rich}_{\text{Correct}} + 0.5) \times (\text{Lean}_{\text{Correct}} + 0.5)}{(\text{Rich}_{\text{Incorrect}} + 0.5) \times (\text{Lean}_{\text{Incorrect}} + 0.5)} \right)$$

High discriminability values are produced by high numbers of correct responses for both rich and lean trials (0.5 is added to all parameters in both equations to avoid instances where no errors are made on a given trial type, thus making log transforms impossible.) The utility of these equations has been repeatedly confirmed in prior studies in humans [3, 4, 8, 29, 32–34] and laboratory animals [2, 9, 11–13, 35]. In addition, accuracy (percent correct calculated by dividing the number of correct trials of each rich/lean trial by the total number of trials of that type) and reaction time (latency from line presentation to response) were calculated and presented as individual subject values and session-wide group means (±SEM) for rich and lean trials.

Electrode implantation

Once the rats acquired line length discrimination, they underwent stereotaxic surgery. Rats were anesthetized with isoflurane (induction 4%, maintenance 2%) in O₂ and secured to a stereotaxic frame with blunted ear bars. Next, rats were surgically implanted with wires in the anterior cingulate cortex (ACC) (AP: +1.2, ML: +0.8, DV: −3.0; all electrodes

were positioned in Cg2) and NAc (AP: +1.2, ML: +0.8, DV: −7.0) for local field potential (LFP) recordings, and two skull screw electrodes near the cerebellum as ground and reference channels. In order to control the distribution of the signal across channels, one of the rats (male) was implanted with two additional skull screw electrodes in the left hemisphere at a frontal and parietal site (Supplementary Fig. 1). The electrodes were connected to an EIB-16 electrode interface board (Neuralynx, Bozeman, MT) and the assembly was secured to the skull using dental acrylic. After a 7-day recovery period, PRT performance was re-established to criteria for the EEG recordings (RHD-2000, Intan Technologies, Los Angeles, CA). At the end of the experiments, the animals were euthanized, and the location of the electrodes was confirmed with Nissl histology (Fig. 1b).

In vivo electrophysiology and data acquisition/reduction

Continuous LFP recordings were acquired during the PRT test session using the RHD-2000 recording system and supported data acquisition software (Intan Technologies, Los Angeles, CA). Signals were locally digitized via a 16-channel headstage and continuously sampled at 1 kHz with a bandwidth range of 0.1–300 Hz for the duration of the behavioral session. Signal analysis was performed using BrainVision Analyzer 2.0 (Brain Products, Gilching, Germany). Data were bandpass filtered between 0.1 and 30 Hz and referenced offline to a common screw electrode placed in the cerebellum. Artifacts were automatically rejected using a minimal-maximal allowed amplitude of −300 μV and 300 μV, respectively, 200 ms before and after the event. Only trials that passed the artifact rejection process were included in these analyses.

Feedback-locked ERP data were segmented into individual epochs spanning from 500 ms before and 1000 ms after feedback, baseline-corrected using a −500 to 0 ms pre-feedback time window and subsequently averaged. Feedback-locked ERPs were quantified as the average activity within the 100–200 ms and 250–500 ms time window following feedback on correct trials only.

For the spectral analysis, the pre-processing steps were similar to those completed for the ERP analysis. Following artifact rejection with the parameters specified above, a complex Morlet wavelet transformation was implemented using a Morlet parameter *c* of 3.5 applied to all trials from 1 to 30 Hz in 30 frequency steps distributed on a logarithmic scale. A percentage change baseline correction was implemented in Analyzer 2.0 (BrainVision Analyzer 2.0 Solution by Dr. Ingmar Gutberlet) by first averaging the amplitude in a 500 to 0 ms pre-feedback time window as the reference interval for this correction. Power values were calculated according to the percentage change in power relative to those baseline periods using the following formula: percentage change (time-frequency) = activity (time-frequency − baseline frequency)/baseline frequency. This percentage change calculation was performed on a trial-by-trial basis. As such, the power values analyzed represent the percentage change in power relative to the baseline period. Subsequently, we applied the average transformation in Analyzer 2.0 and then extracted the wavelet layers of correct trials corresponding to delta (1–5 Hz) at 200–600 ms and 9–17 Hz at 100–200 ms, as an increased activity in the latter frequency band was observed in preliminary analyses. Finally, we created wavelet-decomposed time-frequency spectrograms using the grand average difference in power between rewarded and non-rewarded trials (rewarded − non-rewarded) for each stimulus type; thus, warmer colors represent higher power for rewarded trials.

Statistical analysis

Normal distribution of all data was first tested using the Shapiro–Wilk test and Mauchly's test for sphericity. To analyze response bias and discriminability, a one-way repeated measures ANOVA was conducted to compare either log *b* and log *d*, respectively, across three 3 blocks of 100 trials each. To analyze accuracy and reaction time, a two-way ANOVA with Stimulus Type (rich vs. lean) and Block as factors was performed. Significant main effects or interactions were further examined using Šidák's multiple comparisons test.

To analyze possible time-domain differences between conditions, a two-way ANOVA was performed using the ERP amplitude values at 250–500 ms post-feedback on correct trials for all conditions with Reward Feedback (rewarded vs. non-rewarded) and Stimulus Type (rich vs. lean) as factors. We selected this time window because it approximated the FRN/REwP in humans [23] and visual inspection of the ERP waveforms highlighted deflections differentiating between the feedback conditions. Similarly, a second analogous set of analyses was performed on amplitudes averaged between

100–200 ms as the visual inspection of the ERP waveforms also revealed a differentiation in the ERP between rewarded and non-rewarded trials.

To analyze differences at the spectral level, a two-way ANOVA was applied using the correct trials wavelet average spectral power at 200–600 ms for delta (1–5 Hz) and 100–200 ms for the 9–17 Hz frequency band with Reward Feedback and Stimulus Type as factors. These time windows were selected after visual inspection of the grand average wavelets computed for the overall sample. Significant main effects or interactions were further examined using Šidák's multiple comparisons test.

To probe putative behavior-brain relationships, Pearson correlations were run between response bias and EEG/ERP markers showing main effects of the task. A Bonferroni correction was used to correct for four correlations (rich/lean \times rewarded/non-rewarded; $p = 0.05/4 = 0.0125$). In case of significant correlations, we evaluated the specificity of these findings by running hierarchical regression. For example, if a significant correlation emerged between response bias and feedback-related responses to the rewarded rich stimulus, a regression analysis predicting response bias was performed entering (1) feedback-related responses for all other conditions (e.g., rich non-rewarded, lean rewarded, lean non-rewarded trials) in the first step, and (2) responses to rewarded rich stimulus in the second step. Specificity was formally evaluated by considering ΔF and ΔR^2 .

Statistical analyses were performed using Graph Pad Prism 9 (GraphPad Software Inc., La Jolla, CA) and IBM SPSS® version 24 (IBM Corp, New York, NY).

RESULTS

The rat PRT elicited the intended response bias without fluctuations in task difficulty or reaction time throughout the task

As shown in Fig. 2, the expected responsivity-to-reward behavioral profile was observed. Namely, response bias—which reflects the tendency to correctly classify the more frequently rewarded (rich) stimulus and misclassify the less frequently rewarded (lean) stimulus as being the rich stimulus—was observed and increased across successive blocks (log b , Fig. 2a); this pattern emerged in the context of no significant changes in discriminability, which reflects the ability to differentiate between the stimuli and provides a proxy of task difficulty (log d , Fig. 2b). These signal-detection metrics are collectively a product of increases in accuracy for rich trials paired with decreases in accuracy for lean trials (Fig. 2c), as reliably seen in humans [6]. Reaction times remained consistent across trial blocks (Fig. 2d).

For response bias, the repeated measures ANOVA confirmed a main effect of Block ($F[1.59, 15.9] = 17.50$; $p = 0.0002$), with significant differences between block 1 and 2 ($p = 0.01$) as well as 1 and 3 ($p = 0.0019$), but not between block 2 and 3 ($p = 0.07$). For accuracy, the two-way ANOVA revealed a main effect of Stimulus Type (rich/lean) ($F[1,10] = 52.90$; $p < 0.0001$) due to

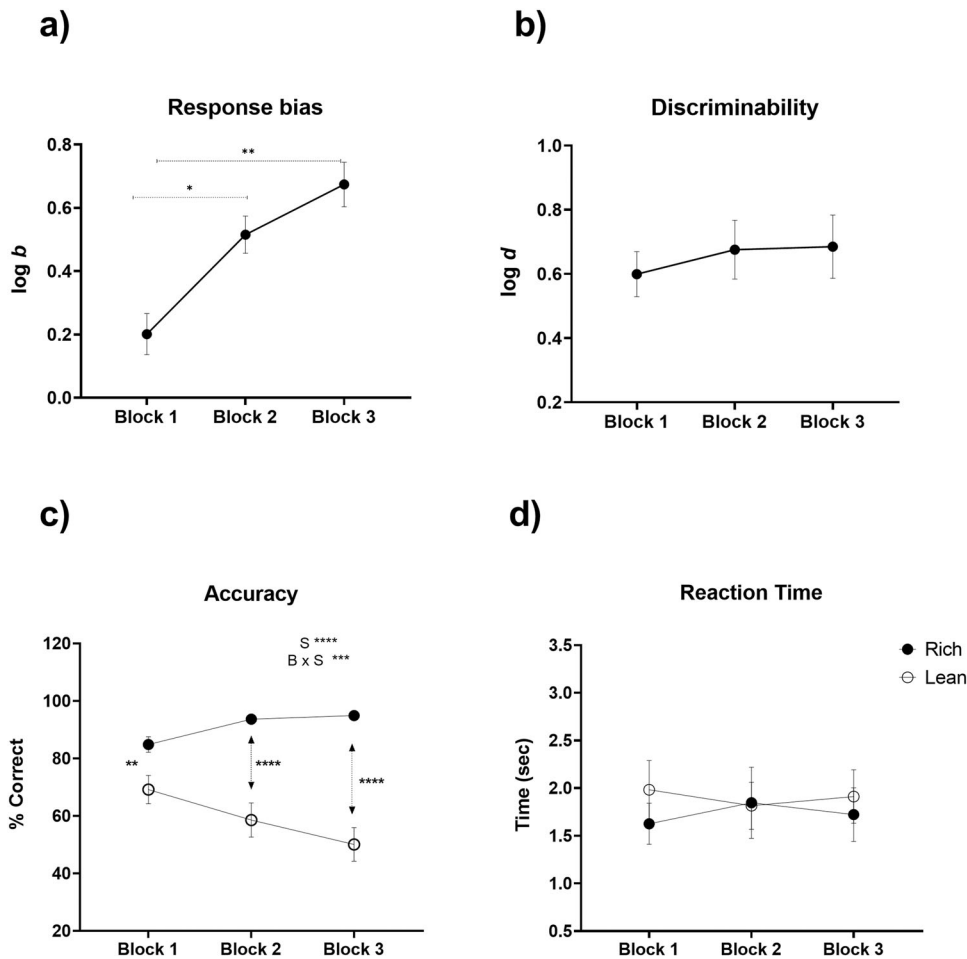


Fig. 2 Behavioral performance in the PRT. The graphs show each variable across 3 blocks of 100 trials each. **a** Response bias, measured with log b , **b** discriminability measured with log d , **c** accuracy, calculated as the percentage of correct responses, and **d** reaction time, measured as time to make a response (seconds). Collectively, these findings indicate that the task evoked the intended preference for the stimulus paired with more frequent reward (log b and higher accuracy for the rich vs. lean stimulus), without fluctuations in task difficulty (log d) or reaction time throughout the task. Main effects and interaction are presented with letters S (Stimulus Type) and B \times S (Block \times Stimulus Type interaction) with asterisks according to their statistical significance. Dotted lines/arrows with asterisks show the statistical significance obtained after multiple comparisons. $N = 11$.

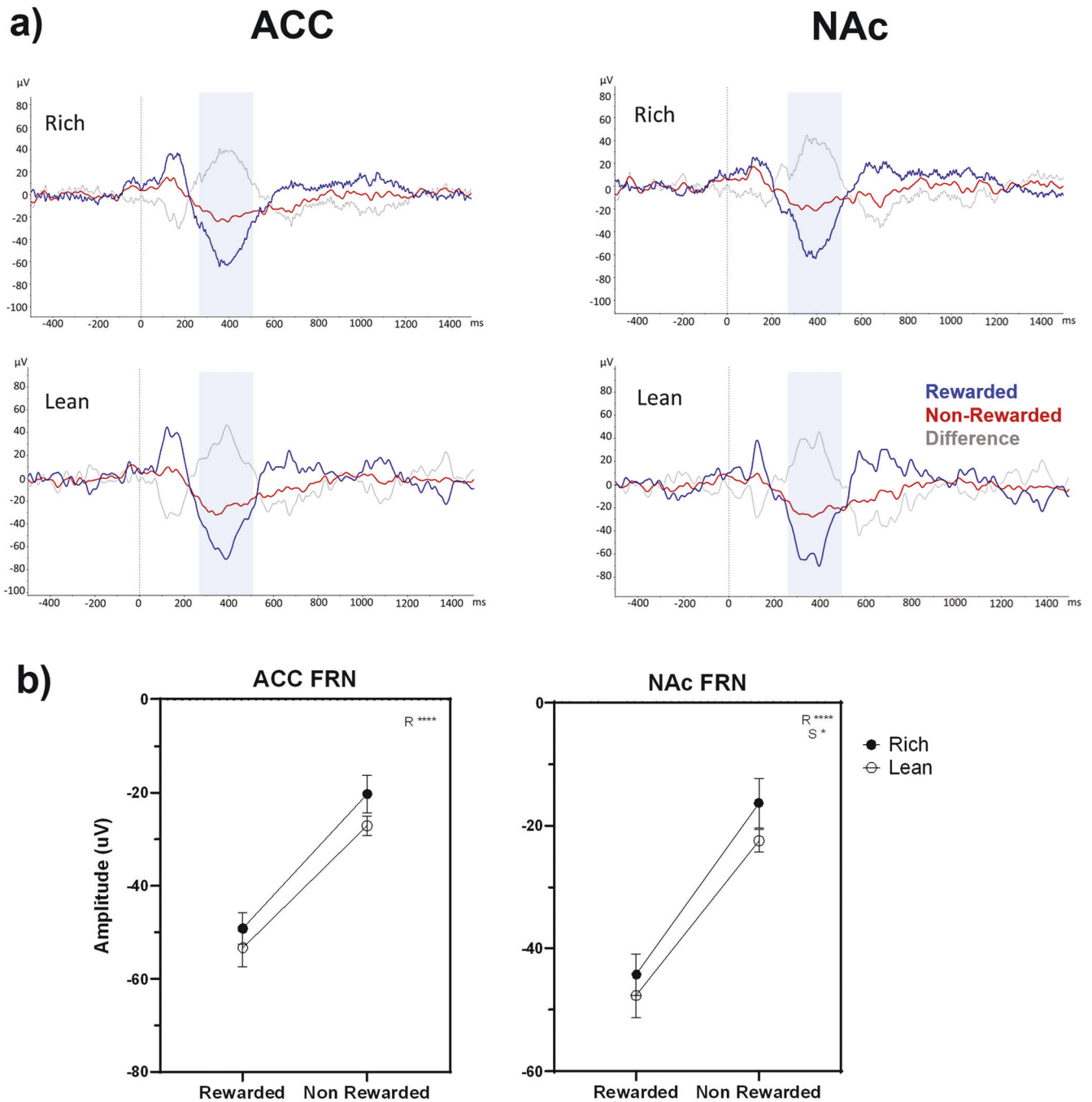


Fig. 3 Feedback-locked event-related potentials (ERPs) in ACC and NAc. **a** Grand average of the feedback-locked ERPs for rewarded (blue), non-rewarded (red) trials, and the difference between them (gray) separated by stimulus type (Left: ACC, right: NAc). A negative deflection at 250–500 ms after feedback similar to an “FRN” (shaded area) was observed in the ACC and NAc local field potentials. Note: The amplitude is presented as microvolts and the time as milliseconds. **b** The graphs show the mean amplitude values in microvolts for the ERPs 250–500 ms after feedback presentation in ACC (left) and NAc (right) for all correct rewarded and non-rewarded trials separated by stimulus type—lean (white circle) and rich (black circle). The magnitudes are presented as mean \pm SEM. Main effects are presented with letters R (reward/non-reward feedback) and S (stimulus type) with asterisks according to their statistical significance $N = 11$.

significantly higher rich than lean accuracy, and a significant Block \times Stimulus Type interaction ($F[2,20] = 11.77$; $p = 0.0004$) due to the difference between rich and lean accuracy becoming larger over blocks (Block 1: $p = 0.005$; Block 2: $p < 0.0001$; Block 3: $p < 0.0001$). Critically, no significant differences in discriminability ($F[1.31, 13.1] = 0.57$; $p > 0.5$) and reaction time ($F[1.937, 19.37] = 0.01$; $p > 0.98$) emerged. Collectively, these findings indicate that the task evoked the intended preference for the stimulus paired with more frequent reward (log b and accuracy),

without fluctuations in task difficulty (log d) or reaction time throughout the task.

The feedback-locked ERP in the rat PRT resembles an “FRN/RewP” waveform in the ACC and NAc local field potentials
In the ERP (Fig. 3a), the first finding was an early positivity occurring around 100–200 ms in the ACC and NAc electrodes; the two-way ANOVA confirmed a main effect of Reward Feedback for the ACC ($F[1,10] = 60.69$; $p < 0.0001$) and NAc ($F[1,10] = 13.53$;

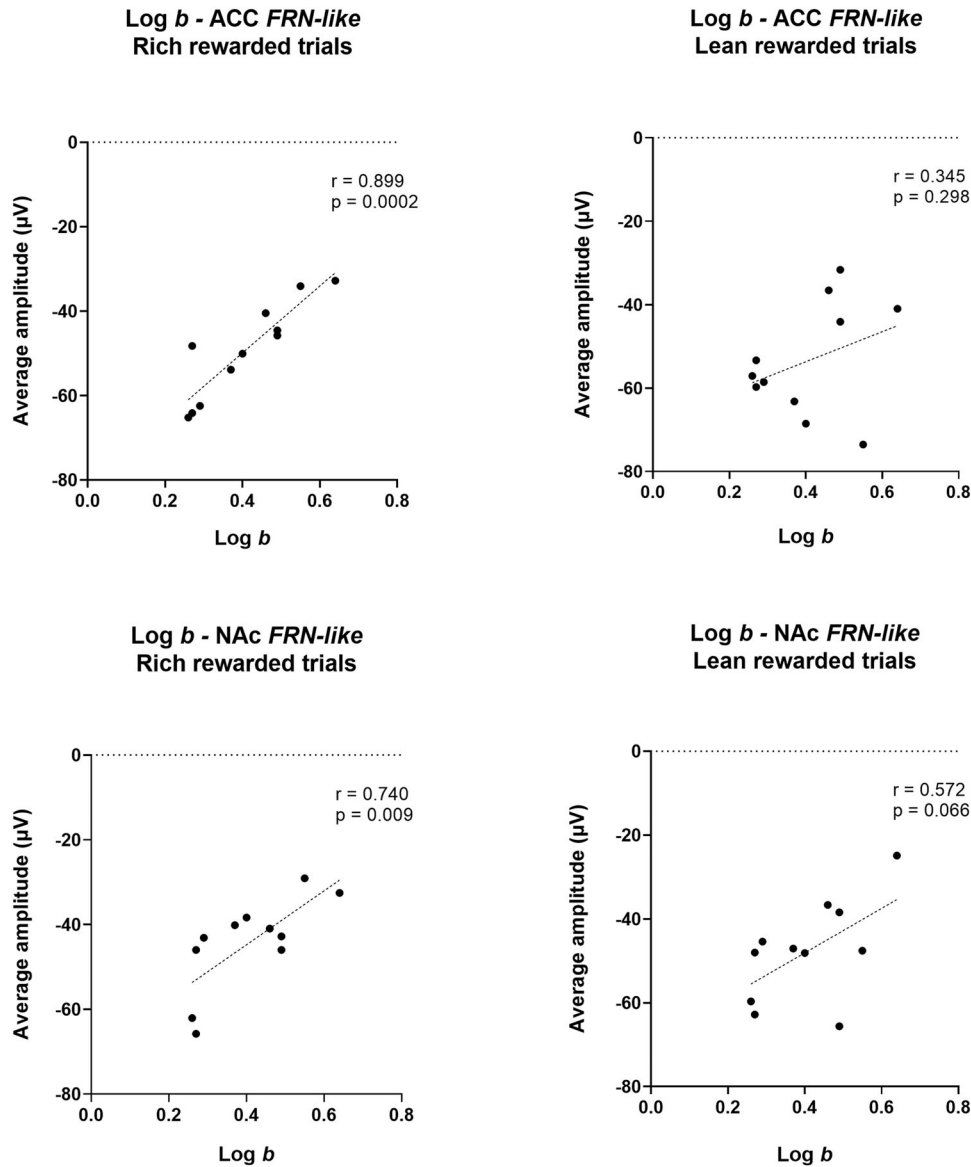


Fig. 4 Correlations between FRN/RewP-like responses and response bias. Scatter plots of the correlations for response bias ($\log b$) and the feedback-locked average amplitude at 250–500 ms (FRN/RewP-like response) for rich rewarded trials in ACC and NAc (top and bottom left, respectively) as well as for lean rewarded trials (top and bottom right). $N = 11$.

$p = 0.004$) due to significantly higher amplitudes for rewarded than non-rewarded trials.

Next, a negative deflection 250–500 ms after reward feedback was observed in the ACC and NAc electrodes, similar to ERP waveforms (FRN/RewP) previously described in humans during the PRT [23], although with an inverse polarity that can be explained by the different type of electrodes used (LFP). As shown in Fig. 3b, for both ACC and NAc channels, this negative deflection was significantly larger for rewarded compared to non-rewarded trials (main effects of Reward Feedback: ACC: $F[1,10] = 132$; $p < 0.0001$; NAc: $F[1,10] = 94.6$; $p < 0.0001$) and a main effect for Stimulus Type (rich vs. lean) also emerged for the NAc ($F[1,10] = 5.65$; $p = 0.04$).

Relationship between FRN/RewP-like responses and response bias

ACC electrode: Pearson correlation analyses (Fig. 4, upper panels) revealed significant positive correlations between response bias and FRN/RewP-like responses to rich rewarded stimuli ($r = 0.899$, $p < 0.0002$; $p < 0.05$ after Bonferroni correction)

and rich non-rewarded stimuli ($r = 0.67$, $p = 0.017$; ns after Bonferroni correction) (all other $ps > 0.28$). A hierarchical regression entering feedback-related responses for rich non-rewarded, lean rewarded and lean non-rewarded trials in the first step, and responses to rewarded rich stimulus in the second step, clarified that FRN/RewP-like ACC responses to rich rewarded stimuli uniquely predicted response bias ($\Delta F[1,6] = 21.18$; $p = 0.004$), accounting for 27.42% of the variance in response bias (variance explained by the other three independent variables: 55.2%, $p > 0.11$). Further highlighting the specificity of these findings, none of the FRN/RewP-like responses correlated with discriminability ($\log d$) ($ps > 0.20$), and an analogous regression considering discriminability was not significant ($p > 0.58$). Collectively, these control analyses indicate that ACC FRN/RewP-like responses to rewarded rich stimuli was specifically associated with the animals' ability to modulate behavior as a function of the probability of reinforcement, rather than general responses to any stimuli or a general ability to discriminate between the stimuli. Critically, in both rats

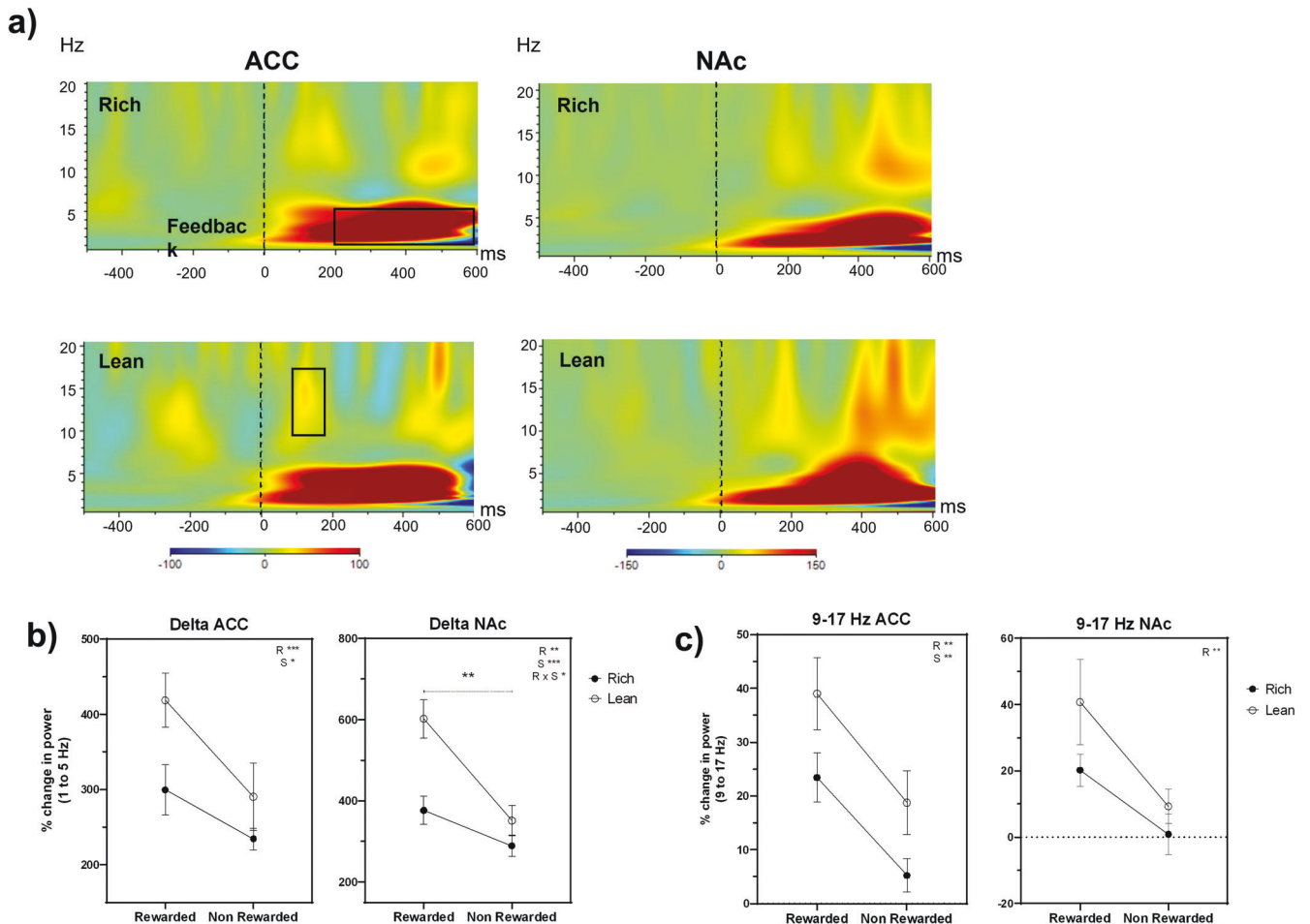


Fig. 5 ACC and NAc feedback-locked wavelet-decomposed time-frequency spectra for correct rich/lean trials. **a** Grand average of feedback-locked wavelet-decomposed time-frequency spectra for correct rich/lean trials presented as the difference between rewarded minus non-rewarded trials for the ACC and NAc channels. The black squares show the time window selected to analyze power difference in delta (1–5 Hz) and 9–17 Hz. **b** The graphs show the mean delta power at 200–600 ms and **c** the mean 9–17 Hz power at 100–200 ms after feedback in ACC and NAc for all correct rewarded and non-rewarded trials separated by stimulus type—lean (white circle) and rich (black circle). The feedback-locked spectral analysis in ACC and NAc revealed increased delta (1–5 Hz) and 9–17 Hz power for rewarded trials. The magnitudes are presented as the percentage change in power from baseline. Data presented as mean \pm SEM. Main effects and interaction are presented with letters R (reward/non-reward feedback), S (stimulus) and R \times S (interaction) with asterisks according to their statistical significance. Dotted lines with asterisks show the statistical significance obtained after multiple comparisons. $N = 11$.

(present study) and humans [23], a significant positive correlation between FRN/RewP and response bias emerged.

NAc electrode: Similar to the ACC, significant positive correlations (Fig. 4, lower panels) emerged between response bias and FRN/RewP-like responses to the rich rewarded stimuli ($r = 0.74$; $p < 0.009$; $p < 0.05$ after Bonferroni correction) and non-rewarded rich stimuli ($r = 0.62$; $p = 0.042$; ns after Bonferroni correction) in the NAc. However, an analogous hierarchical regression analysis for the NAc—entering FRN/RewP-like responses for rich non-rewarded, lean rewarded and lean non-rewarded trials in the first step, and responses to rewarded rich stimulus in the second step—was not significant ($\Delta F[1,6] = 3.55$; $p > 0.10$).

The feedback-locked spectral analysis in ACC and NAc revealed increased delta (1–5 Hz) and 9–17 Hz power for rewarded trials

Feedback-locked 1–5 Hz frequency band. Consistent with human EEG studies highlighting increased delta power in response to rewards and favorable outcomes [22, 36–38], the wavelet frequency decomposition for correct responses highlighted an increased delta power (1–5 Hz) 200–600 ms after reward feedback

in ACC and NAc LFP electrodes (Fig. 5a); these effects were corroborated by main effects of Reward Feedback for both electrodes (ACC: $F[1,10] = 26.60$; $p = 0.0004$; NAc: $F[1,10] = 16.10$; $p = 0.003$), which was driven by significantly higher overall delta power for rewarded than non-rewarded trials (Fig. 5b). Similarly, a main effect of Stimulus Type was found in the ACC and NAc (ACC: $F[1,10] = 5.99$; $p = 0.034$; NAc: $F[1,10] = 24.10$; $p = 0.0006$) as the lean stimulus (which, per design, is associated with the largest positive reward prediction error due to the less frequent rewards) was associated with higher delta power. An interaction between Stimulus Type and Reward Feedback in the NAc was also found ($F[1,10] = 5.98$; $p = 0.035$), due to significantly higher delta power for rewarded vs. non-rewarded lean stimuli ($p = 0.0007$), with no differences for the rich stimulus (Fig. 5b).

Relationship between feedback-locked delta power and response bias
ACC electrode: Pearson correlations revealed a significant association between response bias and ACC delta power for lean rewarded ($r = 0.673$, $p = 0.023$; $p < 0.05$ after Bonferroni correction) (all others $ps > 0.08$). The hierarchical regression (first step: lean non-rewarded, rich rewarded and rich non-rewarded; second step: lean rewarded) revealed that, although delta power to lean rewarded stimuli

accounted for 21.3% unique variance in response bias, the model was not significant ($\Delta F[1,6] = 2.93; p > 0.13$). Of note, the model was significant when controlling only for delta power responses to rich rewarded stimuli in the first step and introducing lean rewarded in the second step, highlighting some degree of specificity ($\Delta F[1,6] = 8.99; p = 0.017$), with delta responses to lean rewarded stimuli accounting for 50.6% unique variance in response bias. Further highlighting specificity, an analogous hierarchical regression predicting discriminability (log d) was not significant ($\Delta F[1,6] = 4.50; p > 0.067$).

NAC electrode: For the NAC, a significant correlation between delta power for lean non-rewarded stimuli and response bias emerged ($r = 0.84; p = 0.001$) (all other $ps > 0.55$). A hierarchical regression (with lean rewarded, rich rewarded and rich non-rewarded in the first step and adding lean non-rewarded in the second step) clarified that delta power in the NAC to lean non-rewarded stimuli uniquely predicted response bias ($\Delta F[1,6] = 21.18; p = 0.004$), accounting for 73.2% of the variance in response bias (variance explained by the other three independent variables: 6.1%, $p > 0.92$). Further highlighting specificity, none of the correlations with discriminability were significant (all $ps > 0.30$), and an analogous regression was not significant ($p > 0.31$).

Feedback-locked 9–17 Hz frequency band. The wavelet frequency decomposition analysis also revealed an increase in 9–17 Hz power band at 100–200 ms after feedback for the ACC and NAC electrodes (Fig. 5a), and the ANOVA confirmed a main effect of Reward Feedback (ACC: $F[1,10] = 13.40; p = 0.004$; NAC: $F[1,10] = 11.20; p = 0.007$) as rewarded trials showed higher 9–17 Hz power than non-rewarded trials. A main effect of Stimulus Type for ACC was also found ($F[1,10] = 12.80; p = 0.005$) as lean trials elicited higher overall 9–17 Hz power than rich trials. A similar trend was observed for the NAC, although it was not statistically significant ($F[1,10] = 3.82; p = 0.079$) (Fig. 5c).

Relationship between feedback-locked 9–17 Hz power and response bias

ACC electrode: Pearson correlations revealed only a significant correlation between response bias and 9–17 Hz power to lean rewarded stimuli ($r = 0.796, p = 0.003; p < 0.05$ after Bonferroni correction) (all others, $ps > 0.16$). A hierarchical regression (with lean non-rewarded, rich rewarded and rich non-rewarded in the first step and adding lean rewarded in the second step) clarified that 9–17 Hz power in the ACC to lean rewarded stimuli uniquely predicted response bias ($\Delta F[1,6] = 6.68; p = 0.042$), accounting for 26.3% of the variance in response bias (variance explained by the other three independent variables: 50.2%, $p > 0.15$). Further highlighting the specificity of these findings, an analogous hierarchical regression predicting discriminability was not significant ($p > 0.68$).

NAC electrode: Pearson correlations revealed no relationship between response bias and any 9–17 Hz power responses in the NAC (all $ps > 0.68$).

DISCUSSION

The overarching goal of the present study was to identify the electrophysiological signatures of reward learning using both ERP and event-related spectral analyses while rats performed a reverse translated task that has been widely used in humans (PRT). Overall, we found that the feedback-locked FRN/RewP-like response, as well as power in feedback-locked delta and 9–17 Hz frequency bands are robust neural signatures of reward learning.

Behaviorally, the expected responsivity-to-reward was observed, confirming previous findings in rodents and highlighting the robustness and reliability of response bias toward the

more frequently rewarded stimulus [2, 9]. Of note, the PRT evoked the intended preference for the stimulus paired with more frequent reward (log b), without fluctuations in task difficulty (log d) or reaction time throughout the task (Fig. 2).

The ERP analyses showed that following a positivity occurring 100–200 ms after feedback in the ACC and NAC for rewarded trials, there was a negative deflection around 250–500 ms in both electrodes (Fig. 3) that we named “FRN/RewP-like” response because it resembles ERP waveforms previously described in humans during the PRT [23, 36]. However, in our LFP experiments we observed a reverse polarity compared to what has been described using surface electrodes in humans. This reverse polarity could be explained by the fact that we recorded deep brain regions, and it has been described that the polarity recorded in LFP is unreliable as it may be artifactually reversed by the zeroing effect of AC-coupled recordings [37]. Additionally, whether an ERP effect is observed to be positive-going or negative-going depends on a variety of factors, such as the location of the reference electrode, the baseline against which the effect is compared, the location and orientation of its intracerebral sources as well as the orientation of the electromagnetic field [38]. This is even more complex to interpret when recording LFP as the polarity does not define their excitatory or inhibitory nature, and the amplitude may increase when source’s activity is reduced [37]. Thus, in the absence of detailed information about the underlying neural activity, the polarity of the ERP effect is of no particular neurophysiological or functional significance for our interpretation.

Critically, and in line with previous studies showing that the ACC is key for encoding rewards and using reinforcement histories to guide behavior [39–42], we found that the ACC FRN/RewP-like responses to rewarded rich stimuli was specifically associated with response bias (Fig. 4). Highlighting important translational potential, it is interesting that in both rats (present study) and humans [23], response bias was positively correlated with FRN/RewP amplitudes.

The time-frequency wavelet decomposition also revealed a finding consistent with human EEG studies implicating delta power in reward processing [25, 43, 44] and in particular with reward prediction error [25]. Specifically, we observed increased delta power 200–600 ms after reward feedback in the ACC and NAC LFP electrodes (Fig. 5) with significantly higher overall power for rewarded than non-rewarded trials, particularly for the lean stimulus, which was supported by a strong correlation between response bias and ACC delta power for lean rewarded trials. This is consistent with the idea that feedback-locked delta emerged in response to a positive prediction error (i.e., when the rats receive a reward that was better than predicted), possibly highlighting an electrophysiological signature of increased phasic neuronal activity of midbrain dopaminergic neurons supporting adaptive learning [15–17]. On the other hand, the finding of a significant correlation between delta power in the NAC for lean non-rewarded stimuli and response bias, may suggest that NAC and ACC delta band differentially modulate response bias depending on the feedback-valence for the lean stimulus (i.e., delta ACC informing about reward vs. delta NAC informing about non-reward); future studies, especially those utilizing more mechanistic (e.g., optogenetic, chemogenetic) approaches, will be needed for conclusive tests of these hypotheses.

In addition, we observed a reward feedback-locked power increase in the 9–17 Hz frequency band at 100–200 ms for the ACC and NAC electrodes (Fig. 5). Similar to the delta findings, 9–17 Hz power in the ACC to lean rewarded stimuli was reliably associated with response bias. We consider this frequency band as a beta rhythm (12–30 Hz) possibly overlapping with some alpha activity (7–12 Hz). It is likely that this 9–17 Hz time-frequency response is a reflection of the 100–200 ms feedback-locked ERP occurring before the FRN/RewP and extends into the delta burst, where

this 9–17 Hz band is part of a larger covarying reward-related complex. Another interpretation is based on evidence showing that, in humans, alpha and beta oscillations during reinforcement learning are associated with the evaluation of reinforcement magnitudes and their subsequent adaptation of response parameters based on this evaluation [45]. Further, our findings are consistent with previous studies suggesting that beta oscillations are stronger in power when the probability of cued rewards have a low—relative to a high—probability [27]—in our case, the lean rewarded trials. Again, more mechanistic studies will be needed to corroborate this speculation.

Despite several innovative features, the current study has limitations. First, although our sample was balanced with respect to male and female subjects, it was not large enough to include sex as a factor in the analyses. As the effect of sex on reward processing is still poorly understood, we believe that preclinical sex comparisons should be considered in the future to better understand the neurophysiological correlates of reward learning. Second, in this initial experiment, we recorded signals from a limited number of LFP channels (but see Supplementary Material for preliminary evidence of topographical specificity of our findings) and we did not use any technique to directly manipulate the circuitry and/or behaviors. Although we offer speculations at the circuit level, and we identified strong brain-behavior associations, the current results are mainly descriptive, and further experimental manipulations should be performed to understand the neural mechanisms that regulate reward learning.

In spite of these limitations, the current data highlight neural signatures of reward learning using electrophysiology and touchscreen-based technology in rats and raise the possibility that rats and humans share similar neural mechanisms to regulate reward learning. It is our hope that the current findings and back-translated approach will provide a useful translational platform to efficiently evaluate novel therapeutics targeting anhedonia.

REFERENCES

- Ang YS, Cusin C, Petibon Y, Dillon DG, Breiger M, Belleau EL, et al. A multi-pronged investigation of option generation using depression, PET and modafinil. *Brain*. 2022;145:1854–65.
- Kangas BD, Der-Avakian A, Pizzagalli DA. Probabilistic reinforcement learning and anhedonia. *Curr Top Behav Neurosci*. 2022;58:355–77.
- Vrieze E, Pizzagalli DA, Demyttenaere K, Hompes T, Sienaert P, de Boer P, et al. Reduced reward learning predicts outcome in major depressive disorder. *Biol Psychiatry*. 2013;73:639–45.
- Pizzagalli DA, Jahn AL, O'Shea JP. Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biol Psychiatry*. 2005;57:319–27.
- National Institute of Mental Health. Behavioral assessment methods for RDoC constructs. 2016.
- Luc OT, Pizzagalli DA, Kangas BD. Toward a quantification of anhedonia: unified matching law and signal detection for clinical assessment and drug development. *Perspect Behav Sci*. 2021;44:517–40.
- Green IW, Pizzagalli DA, Admon R, Kumar P. Anhedonia modulates the effects of positive mood induction on reward-related brain activation. *Neuroimage*. 2019;193:115–25.
- Pizzagalli DA, Goetz E, Ostacher M, Iosifescu DV, Perlis RH. Euthymic patients with bipolar disorder show decreased reward learning in a probabilistic reward task. *Biol Psychiatry*. 2008;64:162–8.
- Kangas BD, Wooldridge LM, Luc OT, Bergman J, Pizzagalli DA. Empirical validation of a touchscreen probabilistic reward task in rats. *Transl Psychiatry*. 2020;10:285.
- Kangas BD, Iturra-Mena AM, Robble MA, Luc OT, Potter D, Nickels S, et al. Concurrent electrophysiological recording and cognitive testing in a rodent touchscreen environment. *Sci Rep*. 2021;11:11665.
- Lamontagne SJ, Melendez SJ, Olmstead MC. Investigating dopamine and glucocorticoid systems as underlying mechanisms of anhedonia. *Psychopharmacology*. 2018;235:3103–13.
- Lamontagne SJ, Wash S, Irwin SH, Zucconi KE, Olmstead MC. Effects of dopamine modulation on chronic stress-induced deficits in reward learning. *Cogn Affect Behav Neurosci*. 2022;22:736–53.
- Wooldridge LM, Bergman J, Pizzagalli DA, Kangas BD. Translational assessments of reward responsiveness in the marmoset. *Int J Neuropsychopharmacol*. 2021;24:409–18.
- Cox J, Witten IB. Striatal circuits for reward learning and decision-making. *Nat Rev Neurosci*. 2019;20:482–94.
- Montague PR, Hyman SE, Cohen JD. Computational roles for dopamine in behavioural control. *Nature*. 2004;431:760–7.
- Schultz W. Multiple dopamine functions at different time courses. *Annu Rev Neurosci*. 2007;30:259–88.
- Kim HR, Malik AN, Mikhael JG, Bech P, Tsutsui-Kimura I, Sun F, et al. A unified framework for dopamine signals across timescales. *Cell*. 2020;183:1600.
- Miltner WHR, Braun CH, Coles MGH. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a “generic” neural system for error detection. *J Cogn Neurosci*. 1997;9:788–98.
- Holroyd CB, Pakzad-Vaezi KL, Krigolson OE. The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*. 2008;45:688–97.
- Proudfit GH. The reward positivity: from basic research on reward to a biomarker for depression. *Psychophysiology*. 2015;52:449–59.
- Krigolson OE. Event-related brain potentials and the study of reward processing: methodological considerations. *Int J Psychophysiol*. 2018;132:175–83.
- Carlson JM, Foti D, Mujica-Parodi LR, Harmon-Jones E, Hajcak G. Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: a combined ERP and fMRI study. *Neuroimage*. 2011;57:1608–16.
- Santesso DL, Dillon DG, Birk JL, Holmes AJ, Goetz E, Bogdan R, et al. Individual differences in reinforcement learning: behavioral, electrophysiological, and neuroimaging correlates. *Neuroimage*. 2008;42:807–16.
- Santesso DL, Ewins AE, Frank MJ, Schetter EC, Bogdan R, Pizzagalli DA. Single dose of a dopamine agonist impairs reinforcement learning in humans: evidence from event-related potentials and computational modeling of striatal-cortical function. *Hum Brain Mapp*. 2009;30:1963–76.
- Cavanagh JF. Cortical delta activity reflects reward prediction error and related behavioral adjustments, but at different times. *Neuroimage*. 2015;110:205–16.
- Marco-Pallares J, Munte TF, Rodriguez-Fornells A. The role of high-frequency oscillatory activity in reward processing and learning. *Neurosci Biobehav Rev*. 2015;49:1–7.
- HajjiHosseini A, Rodriguez-Fornells A, Marco-Pallares J. The role of beta-gamma oscillations in unexpected rewards processing. *Neuroimage*. 2012;60:1678–85.
- Kangas BD, Branch MN. Empirical validation of a procedure to correct position and stimulus biases in matching-to-sample. *J Exp Anal Behav*. 2008;90:103–12.
- Pizzagalli DA, Iosifescu D, Hallett LA, Ratner KG, Fava M. Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. *J Psychiatr Res*. 2008;43:76–87.
- Macmillan NA, Creelman CD. Detection theory: a user's guide. xv. Cambridge: Cambridge University Press; 1990. p. 407.
- Mccarthy D, Davison M. Signal probability, reinforcement and signal-detection. *J Exp Anal Behav*. 1979;32:373–86.
- Boger KD, Auerbach RP, Pechtel P, Busch AB, Greenfield SF, Pizzagalli DA. Co-occurring depressive and substance use disorders in adolescents: an examination of reward responsiveness during treatment. *J Psychother Integr*. 2014;24:109–21.
- Fletcher K, Parker G, Paterson A, Fava M, Iosifescu D, Pizzagalli DA. Anhedonia in melancholic and non-melancholic depressive disorders. *J Affect Disord*. 2015;184:81–8.
- Janes AC, Pedrelli P, Whitton AE, Pechtel P, Douglas S, Martinson MA, et al. Reward responsiveness varies by smoking status in women with a history of major depressive disorder. *Neuropsychopharmacology*. 2015;40:1940–6.
- Der-Avakian A, D'Souza MS, Pizzagalli DA, Markou A. Assessment of reward responsiveness in the response bias probabilistic reward task in rats: implications for cross-species translational research. *Transl Psychiatry*. 2013;3:e297.
- Bogdan R, Santesso DL, Fagerness J, Perlis RH, Pizzagalli DA. Corticotropin-releasing hormone receptor type 1 (CRHR1) genetic variation and stress interact to influence reward learning. *J Neurosci*. 2011;31:13246–54.
- Herreras O. Local field potentials: myths and misunderstandings. *Front Neural Circuits*. 2016;10:101.
- Picton TW. Electrophysiology of mind: event-related brain potentials and cognition—Rugg,MD, Coles,MGH. *Psychophysiology*. 1996;33:612–3.
- Akitsuki Y, Sugiura M, Watanabe J, Yamashita K, Sassa Y, Awata S, et al. Context-dependent cortical activation in response to financial reward and penalty: an event-related fMRI study. *Neuroimage*. 2003;19:1674–85.
- Amiez C, Joseph JP, Procyk E. Reward encoding in the monkey anterior cingulate cortex. *Cereb Cortex*. 2006;16:1040–55.
- Smith BW, Mitchell DG, Hardin MG, Jazbec S, Fridberg D, Blair RJ, et al. Choice selection and reward anticipation: an fMRI study. *Neuropsychologia*. 2004;42:1585–97.

42. Rushworth MF, Buckley MJ, Behrens TE, Walton ME, Bannerman DM. Functional organization of the medial frontal cortex. *Curr Opin Neurobiol.* 2007;17:220–7.
43. Foti D, Weinberg A, Bernat EM, Proudfit GH. Anterior cingulate activity to monetary loss and basal ganglia activity to monetary gain uniquely contribute to the feedback negativity. *Clin Neurophysiol.* 2015;126:1338–47.
44. Cavanagh JF, Olguin SL, Talledo JA, Kotz JE, Roberts BZ, Nungaray JA, et al. Amphetamine alters an EEG marker of reward processing in humans and mice. *Psychopharmacology.* 2022;239:923–33.
45. Schroll H, Horn A, Runge J, Lipp A, Schneider GH, Krauss JK, et al. Reinforcement magnitudes modulate subthalamic beta band activity in patients with Parkinson's disease. *Sci Rep.* 2018;8:8621.

AUTHOR CONTRIBUTIONS

AMI, BDK and DAP participated in the research design. AMI and OTL conducted the experiments. DP performed the histology. AMI, BDK and DAP performed data analysis. AMI, BDK and DAP contributed substantially to the writing of this manuscript. DAP secured funding. All authors have approved the final version of this manuscript.

FUNDING

This work was funded by 1 P50 MH119467-02 (to DAP).

COMPETING INTERESTS

Over the past 3 years, DAP has received consulting fees from Albright Stonebridge Group, Boehringer Ingelheim, Compass Pathways, Engrail Therapeutics, Neumora Therapeutics (formerly BlackThorn Therapeutics), Neurocrine Biosciences, Neuroscience Software, Otsuka, Sunovion, and Takeda; he has received honoraria from

the Psychonomic Society and American Psychological Association (for editorial work) and from Alkermes; he has received research funding from the Brain and Behavior Research Foundation, Dana Foundation, Wellcome Leap, Millennium Pharmaceuticals, and NIMH; he has received stock options from Compass Pathways, Engrail Therapeutics, Neumora Therapeutics, and Neuroscience Software; he has a financial interest in Neumora Therapeutics, which has licensed the copyright to the human version of the probabilistic reward task through Harvard University. No funding from these entities was used to support the current work, and all views expressed are solely those of the authors. All other authors have no conflicts of interest or relevant disclosures.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41386-023-01532-4>.

Correspondence and requests for materials should be addressed to Diego A. Pizzagalli.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.